

Docket No.: 209013US55CONT

IN THE UNITED STATES PATENT & TRADEMARK OFFICE

IN RE APPLICATION OF: :

John B. LOWE : ATTN: APPLICATION DIVISION

SERIAL NO: NEW APPLICATION :

FILED: HERewith :

FOR: METHODS AND PRODUCTS FOR THE SYNTHESIS OF OLIGOSACCHARIDE
STRUCTURES ON GLYCOPROTEINS, GLYCOLIPIDS, OR AS FREE
MOLECULES, AND FOR THE ISOLATION OF CLONED GENETIC
SEQUENCES THAT DETERMINE THESE STRUCTURES

SEQUENCE LISTING STATEMENT

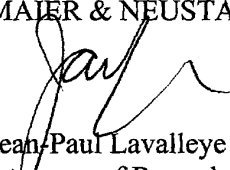
ASSISTANT COMMISSIONER FOR PATENTS
WASHINGTON, D.C. 20231

SIR:

Applicant respectfully request that the computer-readable form in parent application
Serial No. 07/914,281 be used to prepare a Sequence Listing file for the above-identified
application. The contents of the paper copy of the Sequence Listing in the above-identified
application and the computer-readable form in parent application Serial No. 07/914,281 are
identical.

Respectfully submitted,

OBLON, SPIVAK, McCLELLAND,
MAIER & NEUSTADT, P.C.


Jean-Paul Lavalleye
Attorney of Record
Registration No. 31,451

James J. Kelly, Ph.D.
Registration No. 41,504



22850

Tel. (703) 413-3000
Fax. (703) 413-2220
(OSMMN 11/98)

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- (i) APPLICANT: LOWE, JOHN B.
- (ii) TITLE OF INVENTION: METHODS AND PRODUCTS FOR THE SYNTHESIS OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS, GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
- (iii) NUMBER OF SEQUENCES: 14
- (iv) CORRESPONDENCE ADDRESS:
 - (A) ADDRESSEE: OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT, P.C.
 - (B) STREET: 1755 Jefferson Davis Highway, Fourth Floor
 - (C) CITY: Arlington
 - (D) STATE: Virginia
 - (E) COUNTRY: U.S.A.
 - (F) ZIP: 22202
- (v) COMPUTER READABLE FORM:
 - (A) MEDIUM TYPE: Floppy disk
 - (B) COMPUTER: IBM PC compatible
 - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 - (D) SOFTWARE: PatentIn Release #1.0, Version #1.25
- (vi) CURRENT APPLICATION DATA:
 - (A) APPLICATION NUMBER: US
 - (B) FILING DATE: 20-JUL-1992
 - (C) CLASSIFICATION:
- (viii) ATTORNEY/AGENT INFORMATION:
 - (A) NAME: Lavalleye, Jean-Paul M. P.
 - (B) REGISTRATION NUMBER: 31,451
 - (C) REFERENCE/DOCKET NUMBER: 2363-060-55
- (ix) TELECOMMUNICATION INFORMATION:
 - (A) TELEPHONE: (703)521-4500
 - (B) TELEFAX: (703)486-2347
 - (C) TELEX: 248855 OPAT UR

(2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 2043 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: unknown
 - (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

AGGAAACCTG CCATGGCCTC CTGGTGAGCT GTCCTCATCC ACTGCTCGCT GCCTCTCCAG 60
ATACTCTGAC CCATGGATCC CTTGGGTGCA GCCAAGCCAC AATGGCCATG GCGCCGCTGT 120
CTGGCCGCAC TGCTATTTCA GCTGCTGGTG GCTGTGTGTT TCTTCTCCTA CCTGCGTGTG 180
TCCCGAGACG ATGCCACTGG ATCCCCTAGG GCTCCCAGTG GGTCCCTCCCG ACAGGACACC 240
ACTCCCACCC GCCCCACCCT CTTGATCCTG CTATGGACAT GGCCTTTCCA CATCCCTGTG 300
GCTCTGTCCC GCTGTTTACA GATGGTGCCC GGCACAGCCG ACTGCCACAT CACTGCCGAC 360
CGCAAGGTGT ACCCACAGGC AGACACGGTC ATCGTGCACC ACTGGGATAT CATGTCCAAC 420
CCTAAGTCAC GCCTCCCACC TTCCCCGAGG CCGCAGGGGC AGCGCTGGAT CTGGTTCAAC 480
TTGGAGCCAC CCCCTAACTG CCAGCACCTG GAAGCCCTGG ACAGATACTT CAATCTCACC 540
ATGTCCTACC GCAGCGACTC CGACATCTTC ACGCCCTACG GCTGGCTGGA GCCGTGGTCC 600
GGCCAGCCTG CCCACCCACC GCTCAACCTC TCGGCCAAGA CCGAGCTGGT GGCCTGGGCG 660
GTGTCCAAC GGAAGCCGGA CTCAGCCAGG GTGCGCTACT ACCAGAGCCT GCAGGCTCAT 720
CTCAAGGTGG ACGTGTACGG ACGCTCCCAC AAGCCCCTGC CCAAGGGGAC CATGATGGAG 780
ACGCTGTCCC GGTACAAGTT CTACCTGGCC TTCGAGAACT CCTTGCACCC CGACTACATC 840
ACCGAGAAGC TGTGGAGGAA CGCCCTGGAG GCCTGGGCGG TGCCCGTGGT GCTGGGCCCC 900
AGCAGAAGCA ACTACGAGAG GTTCCTGCCA CCCGACGCCT TCATCCACGT GGACGACTTC 960
CAGAGCCCCA AGGACCTGGC CCGGTACCTG CAGGAGCTGG ACAAGGACCA CGCCCGCTAC 1020
CTGAGCTACT TTCGCTGGCG GGAGACGCTG CGGCCTCGCT CCTTCAGCTG GGCCTGGAT 1080
TTCTGCAAGG CCTGCTGGAA ACTGCAGCAG GAATCCAGGT ACCAGACGGT GCGCAGCATA 1140
GCGGCTTGGT TCACCTGAGA GGCCGGCATG GTGCCTGGGC TGCCGGGAAC CTCATCTGCC 1200
TGGGGCCTCA CCTGCTGGAG TCCTTTGTGG CCAACCCTCT CTCTTACCTG GGACCTCACA 1260
CGCTGGGCTT CACGGCTGCC AGGAGCCTCT CCCCTCCAGA AGACTTGCCT GCTAGGGACC 1320

TCGCCTGCTG GGGACCTCGC CTGTTGGGGA CCTCACCTGC TGGGGACCTC ACCTGCTGGG 1380
 GACCTTGGCT GCTGGAGGCT GCACCTACTG AGGATGTCGG CGGTCGGGGA CTTTACCTGC 1440
 TGGGACCTGC TCCCAGAGAC CTTGCCACAC TGAATCTCAC CTGCTGGGGA CCTCACCTG 1500
 GAGGGCCCTG GGCCCTGGGG AACTGGCTTA CTTGGGGCCC CACCCGGGAG TGATGGTTCT 1560
 GGCTGATTTG TTTGTGATGT TGTTAGCCGC CTGTGAGGGG TGCAGAGAGA TCATCACGGC 1620
 ACGGTTTCCA GATGTAATAC TGCAAGGAAA AATGATGACG TGTCTCCTCA CTCTAGAGGG 1680
 GTTGGTCCCA TGGGTTAAGA GCTCACCCCA GGTTCCTACC TCAGGGGTTA AGAGCTCAGA 1740
 GTTCAGACAG GTCCAAGTTC AAGCCCAGGA CCACCACTTA TAGGGTACAG GTGGGATCGA 1800
 CTGTAAATGA GGA CTCTGG AACATTCCAA ATATTCTGGG GTTGAGGGAA ATTGCTGCTG 1860
 TCTACAAAAT GCCAAGGGTG GACAGGCGCT GTGGCTCACG CCTGTAATTC CAGCACTTTG 1920
 GGAGGCTGAG GTAGGAGGAT TGATTGAGGC CAAGAGTTAA AGACCAGCCT GGTCAATATA 1980
 GCAAGACCAC GTCTCTAAAT AAAAAATAAT AGGCCGGCCA GGAAAAAAAA AAAAAAAAAA 2040
 AAA 204

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 361 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

| | | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Met | Asp | Pro | Leu | Gly | Ala | Ala | Lys | Pro | Gln | Trp | Pro | Trp | Arg | Arg | Cys |
| 1 | | | | 5 | | | | 10 | | | | | | 15 | |
| Leu | Ala | Ala | Leu | Leu | Phe | Gln | Leu | Leu | Val | Ala | Val | Cys | Phe | Phe | Ser |
| | | | 20 | | | | 25 | | | | | | 30 | | |
| Tyr | Leu | Arg | Val | Ser | Arg | Asp | Asp | Ala | Thr | Gly | Ser | Pro | Arg | Ala | Pro |
| | | 35 | | | | 40 | | | | | | 45 | | | |
| Ser | Gly | Ser | Ser | Arg | Gln | Asp | Thr | Thr | Pro | Thr | Arg | Pro | Thr | Leu | Leu |
| | 50 | | | | | 55 | | | | | 60 | | | | |

| | | | | | | | | | | | | | | | |
|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|
| Ile 65 | Leu | Leu | Trp | Thr | Trp 70 | Pro | Phe | His | Ile | Pro 75 | Val | Ala | Leu | Ser | Arg 80 |
| Cys | Ser | Glu | Met | Val 85 | Pro | Gly | Thr | Ala | Asp 90 | Cys | His | Ile | Thr | Ala 95 | Asp |
| Arg | Lys | Val | Tyr 100 | Pro | Gln | Ala | Asp | Thr 105 | Val | Ile | Val | His | His 110 | Trp | Asp |
| Ile | Met | Ser 115 | Asn | Pro | Lys | Ser | Arg 120 | Leu | Pro | Pro | Ser | Pro 125 | Arg | Pro | Gln |
| Gly | Gln 130 | Arg | Trp | Ile | Trp | Phe 135 | Asn | Leu | Glu | Pro | Pro 140 | Pro | Asn | Cys | Gln |
| His 145 | Leu | Glu | Ala | Leu | Asp 150 | Arg | Tyr | Phe | Asn | Leu 155 | Thr | Met | Ser | Tyr | Arg 160 |
| Ser | Asp | Ser | Asp | Ile 165 | Phe | Thr | Pro | Tyr | Gly 170 | Trp | Leu | Glu | Pro | Trp 175 | Ser |
| Gly | Gln | Pro | Ala 180 | His | Pro | Pro | Leu | Asn 185 | Leu | Ser | Ala | Lys | Thr 190 | Glu | Leu |
| Val | Ala | Trp 195 | Ala | Val | Ser | Asn | Trp 200 | Lys | Pro | Asp | Ser | Ala 205 | Arg | Val | Arg |
| Tyr | Tyr 210 | Gln | Ser | Leu | Gln | Ala 215 | His | Leu | Lys | Val | Asp 220 | Val | Tyr | Gly | Arg |
| Ser 225 | His | Lys | Pro | Leu | Pro 230 | Lys | Gly | Thr | Met | Met 235 | Glu | Thr | Leu | Ser | Arg 240 |
| Tyr | Lys | Phe | Tyr | Leu 245 | Ala | Phe | Glu | Asn | Ser 250 | Leu | His | Pro | Asp | Tyr 255 | Ile |
| Thr | Glu | Lys | Leu 260 | Trp | Arg | Asn | Ala | Leu 265 | Glu | Ala | Trp | Ala | Val 270 | Pro | Val |
| Val | Leu | Gly 275 | Pro | Ser | Arg | Ser | Asn 280 | Tyr | Glu | Arg | Phe | Leu 285 | Pro | Pro | Asp |
| Ala | Phe 290 | Ile | His | Val | Asp | Asp 295 | Phe | Gln | Ser | Pro | Lys 300 | Asp | Leu | Ala | Arg |
| Tyr 305 | Leu | Gln | Glu | Leu | Asp 310 | Lys | Asp | His | Ala | Arg 315 | Tyr | Leu | Ser | Tyr | Phe 320 |
| Arg | Trp | Arg | Glu | Thr 325 | Leu | Arg | Pro | Arg | Ser 330 | Phe | Ser | Trp | Ala | Leu | Asp 335 |

Phe Cys Lys Ala Cys Trp Lys Leu Gln Gln Glu Ser Arg Tyr Gln Thr
 340 345 350

Val Arg Ser Ile Ala Ala Trp Phe Thr
 355 360

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1500 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

| | |
|---|-----|
| CCTTCCCTTG TAGACTCTTC TTGGAATGAG AAGTACCGAT TCTGCTGAAG ACCTCGCGCT | 60 |
| CTCAGGCTCT GGGAGTTGGA ACCCTGTACC TTCCTTTCCT CTGCTGAGCC CTGCCTCCTT | 120 |
| AGGCAGGCCA GAGCTCGACA GAACTCGGTT GCTTTGCTGT TTGCTTTGGA GGGAACACAG | 180 |
| CTGACGATGA GGCTGACTTT GAACTCAAGA GATCTGCTTA CCCCAGTCTC CTGGAATTAA | 240 |
| AGGCCTGTAC TACATTGCCC TGGACCTAAG ATTTTCATGA TCACTATGCT TCAAGATCTC | 300 |
| CATGTCAACA AGATCTCCAT GTCAAGATCC AAGTCAGAAA CAAGTCTTCC ATCCTCAAGA | 360 |
| TCTGGATCAC AGGAGAAAAT AATGAATGTC AAGGGAAAAG TAATCCTGTT GATGCTGATT | 420 |
| GTCTCAACCG TGGTTGTCGT GTTTTGGGAA TATGTCAACA GAATTCCAGA GGTGTTGAG | 480 |
| AACAGATGGC AGAAGGACTG GTGGTTCCCA AGCTGGTTTA AAAATGGGAC CCACAGTTAT | 540 |
| CAAGAAGACA ACGTAGAAGG ACGGAGAGAA AAGGGTAGAA ATGGAGATCG CATTGAAGAG | 600 |
| CCTCAGCTAT GGGACTGGTT CAATCCAAAG AACCGCCCGG ATGTTTTGAC AGTGACCCCG | 660 |
| TGGAAGGCGC CGATTGTGTG GGAAGGCACT TATGACACAG CTCTGCTGGA AAAGTACTAC | 720 |
| GCCACACAGA AACTCACTGT GGGGCTGACA GTGTTTGCTG TGGGAAAGTA CATTGAGCAT | 780 |
| TACTTAGAAG ACTTTCTGGA GTCTGCTGAC ATGTACTTCA TGGTTGGCCA TCGGGTCATA | 840 |
| TTTACGTCA TGATAGACGA CACCTCCCGG ATGCCTGTCTG TGCACCTGAA CCCTCTACAT | 900 |

TCCTTACAAG TCTTTGAGAT CAGGTCTGAG AAGAGGTGGC AGGATATCAG CATGATGCGC 960
 ATGAAGACCA TTGGGGAGCA CATCCTGGCC CACATCCAGC ACGAGGTCTGA CTCCTCTTC 1020
 TGCATGGACG TGGATCAAGT CTTTCAAGAC AACTTCGGGG TGGAAACTCT GGGCCAGCTG 1080
 GTAGCACAGC TCCAGGCCTG GTGGTACAAG GCCAGTCCCG AGAAGTTCAC CTATGAGAGG 1140
 CGGGAAGTGT CGGCCGCGTA CATTCCATTC GGAGAGGGGG ATTTTACTA CCACGCGGCC 1200
 ATTTTGGAG GAACGCCTAC TCACATTCTC AACCTCACCA GGGAGTGCTT TAAGGGGATC 1260
 CTCCAGGACA AGAAACATGA CATAGAAGCC CAGTGGCATG ATGAGAGCCA CCTCAACAAA 1320
 TACTTCCTTT TCAACAAACC CACTAAAATC CTATCTCCAG AGTATTGCTG GGACTATCAG 1380
 ATAGGCCTGC CTTCAGATAT TAAAGTGTC AAGGTAGCTT GGCAGACAAA AGAGTATAAT 1440
 TTGGTTAGAA ATAATGTCTG ACTTCAAATT GTGATGGAAA CTTGACACTA TTTCTAACCA 1500

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 394 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

| | | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Met | Ile | Thr | Met | Leu | Gln | Asp | Leu | His | Val | Asn | Lys | Ile | Ser | Met | Ser |
| 1 | | | | 5 | | | | | 10 | | | | | 15 | |
| Arg | Ser | Lys | Ser | Glu | Thr | Ser | Leu | Pro | Ser | Ser | Arg | Ser | Gly | Ser | Gln |
| | | | 20 | | | | | 25 | | | | | 30 | | |
| Glu | Lys | Ile | Met | Asn | Val | Lys | Gly | Lys | Val | Ile | Leu | Leu | Met | Leu | Ile |
| | | 35 | | | | | 40 | | | | | 45 | | | |
| Val | Ser | Thr | Val | Val | Val | Val | Phe | Trp | Glu | Tyr | Val | Asn | Arg | Ile | Pro |
| | | 50 | | | | 55 | | | | | 60 | | | | |
| Glu | Val | Gly | Glu | Asn | Arg | Trp | Gln | Lys | Asp | Trp | Trp | Phe | Pro | Ser | Trp |
| 65 | | | | | 70 | | | | | 75 | | | | | 80 |
| Phe | Lys | Asn | Gly | Thr | His | Ser | Tyr | Gln | Glu | Asp | Asn | Val | Glu | Gly | Arg |
| | | | | 85 | | | | | 90 | | | | | 95 | |

| | | | | | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Arg | Glu | Lys | Gly | Arg | Asn | Gly | Asp | Arg | Ile | Glu | Glu | Pro | Gln | Leu | Trp | 100 | 105 | 110 |
| Asp | Trp | Phe | Asn | Pro | Lys | Asn | Arg | Pro | Asp | Val | Leu | Thr | Val | Thr | Pro | 115 | 120 | 125 |
| Trp | Lys | Ala | Pro | Ile | Val | Trp | Glu | Gly | Thr | Tyr | Asp | Thr | Ala | Leu | Leu | 130 | 135 | 140 |
| Glu | Lys | Tyr | Tyr | Ala | Thr | Gln | Lys | Leu | Thr | Val | Gly | Leu | Thr | Val | Phe | 145 | 150 | 155 |
| Ala | Val | Gly | Lys | Tyr | Ile | Glu | His | Tyr | Leu | Glu | Asp | Phe | Leu | Glu | Ser | 165 | 170 | 175 |
| Ala | Asp | Met | Tyr | Phe | Met | Val | Gly | His | Arg | Val | Ile | Phe | Tyr | Val | Met | 180 | 185 | 190 |
| Ile | Asp | Asp | Thr | Ser | Arg | Met | Pro | Val | Val | His | Leu | Asn | Pro | Leu | His | 195 | 200 | 205 |
| Ser | Leu | Gln | Val | Phe | Glu | Ile | Arg | Ser | Glu | Lys | Arg | Trp | Gln | Asp | Ile | 210 | 215 | 220 |
| Ser | Met | Met | Arg | Met | Lys | Thr | Ile | Gly | Glu | His | Ile | Leu | Ala | His | Ile | 225 | 230 | 235 |
| Gln | His | Glu | Val | Asp | Phe | Leu | Phe | Cys | Met | Asp | Val | Asp | Gln | Val | Phe | 245 | 250 | 255 |
| Gln | Asp | Asn | Phe | Gly | Val | Glu | Thr | Leu | Gly | Gln | Leu | Val | Ala | Gln | Leu | 260 | 265 | 270 |
| Gln | Ala | Trp | Trp | Tyr | Lys | Ala | Ser | Pro | Glu | Lys | Phe | Thr | Tyr | Glu | Arg | 275 | 280 | 285 |
| Arg | Glu | Leu | Ser | Ala | Ala | Tyr | Ile | Pro | Phe | Gly | Glu | Gly | Asp | Phe | Tyr | 290 | 295 | 300 |
| Tyr | His | Ala | Ala | Ile | Phe | Gly | Gly | Thr | Pro | Thr | His | Ile | Leu | Asn | Leu | 305 | 310 | 315 |
| Thr | Arg | Glu | Cys | Phe | Lys | Gly | Ile | Leu | Gln | Asp | Lys | Lys | His | Asp | Ile | 325 | 330 | 335 |
| Glu | Ala | Gln | Trp | His | Asp | Glu | Ser | His | Leu | Asn | Lys | Tyr | Phe | Leu | Phe | 340 | 345 | 350 |
| Asn | Lys | Pro | Thr | Lys | Ile | Leu | Ser | Pro | Glu | Tyr | Cys | Trp | Asp | Tyr | Gln | 355 | 360 | 365 |

100 110 120 130 140 150 160 170 180 190 200 210 220 230 240 250 260 270 280 290 300 310 320 330 340 350 360

Ile Gly Leu Pro Ser Asp Ile Lys Ser Val Lys Val Ala Trp Gln Thr
 370 375 380

Lys Glu Tyr Asn Leu Val Arg Asn Asn Val
 385 390

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 8174 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

| | |
|---|-----|
| GAATTCCATC GTGGCAAGGG CAGCCTGAAT GGATGATGTA ACCTGGGGTC CTTTCAATGG | 60 |
| AGGGCCAGAC TCCTGGGTCT AGGGGATGAG GGAGGGGAGG ATCGGGTTAG CTGGGACCCA | 120 |
| GGTGAAAGGG GCTGGGGGCC CACATTCCTG AGTCTCAGAG AGAAGGATCT GGGGTCTCAA | 180 |
| GCACCTGAGT CGGAGGGAGG AGGGGTGCTG GGCTCCTGGA AAAACCACCT CTTGGACCAT | 240 |
| CTATGCAGAT CACGCAGAAC AAGAGAAATT TCTGCGCCCC ATCTGAATTT CTAAGTTTGG | 300 |
| GGGGAGGGCG TGATCTGACA CTGAGGTTCC TTGATCCTCA GCAAGGCGGC AATTGCTGTA | 360 |
| TGAAAGAAGC GACCGCATCT GAGACACAAG TATCCTGCCT TGGAAGCCTC TCACCTGGCC | 420 |
| GTGGGCCAAC CTCAACCTCA TCTGTCCCTG CTCAGATGCT CAGACCCTGG ACATCCCAGC | 480 |
| CTCCTCCTCC CTGATGCAAT CCTGGTGTTT CTTTCACCAG AGAAGCCATC CCAGGCCCAG | 540 |
| GCAGGTGCTC CTGAAATAAC CTGGGGGGAG GGGTGGCTGA AAGTCCCTGA CTGGAGTTGG | 600 |
| CAGCCAAGCC AGGCCCTGGA GTGGGCACCC AGAGGGAAGA CAGGTTGGCT AATTTCCTGG | 660 |
| AGCCCCTAAG GGTGCAAGGG TAGGCCTTCT GTGTCTGAGG GAGGAGGGCT GGGGCTCTGG | 720 |
| ACTCCTGGGT CTGAGGGAGG AGGGGTGGGG GGCCTGGACT CCTGGGTCTG AGGGAGGAGG | 780 |
| GTCTGGGCCT GTACTCCTGG ATCTGAGGGA GGAGGGGCTG GGGAACTTGG GCTCCTGGGT | 840 |
| CTGAGGGAGG AGGGAGCTTT GGTCTGGACT CCTGGGTCTG AGGGAGTAGG GGCTAGGGAT | 900 |

10450-54500

| | |
|---|------|
| CTGGACTCGT GGGTGTGAGG AAGGAGGGGC TGGGGTCCTG GACTCCTGGG TCTGAGGAAG | 960 |
| GAGGGGCAGG GGGCTTGAC TCCTGGGTCT GAGGAAGGAG GGGCCGGGAG CCTGGACTCC | 1020 |
| TAAGTCTGAG GGAGGAGGGT CTGGGGGCCT GGAAGTCTGG GTGTGAGCAG AAGGGTCTGG | 1080 |
| GTGCTGGGAG TCCCGAGCCT GGGGAGATGA TGGTTAAACT TCTGGGAATC AAGTCAAAC | 1140 |
| CCTGAGTCTT TGACATTGAT GTATCTTGAA TGGGAGGGTC AGTCTGTGGG GAAGGATTAC | 1200 |
| CCAGGTGCCG AGGCAAGAGA CTGAAGGCAC AAAGTGTTC AGTATAATAA AGAAAATAGT | 1260 |
| TAGAATAAGA ATAGTTATCA TACAAATTAG ATATAGAGAT GATCATGGAC AGTATCAATC | 1320 |
| ATTAGTGTA ACATTATTAA TCATTAGCTA TTAAGTTTAT TCTTTGTTGT ATAAGTAATA | 1380 |
| TAACCAGGAA ACAACCGGTG GGTATAGGGT CAGGTACTGA AGGGACATTG TGAGAAGTGA | 1440 |
| CCTAGAAGGC AAGAGGTGAG CCTTCTGTCA CACCGGCATA AGGGCCTCTT GAGGGCTCCT | 1500 |
| TGGTCAAGCG GGAACGCCAG TGTCTGGGAA GGCACCGTT ACTCAGCAGA CCACGAAAGG | 1560 |
| GAATCTCCTT TTCTTGGAGG AGTCAGGGAA CACTCTGCTC CACAGCTTC TTGTGGGAGG | 1620 |
| CTGGGTATTA TCTAGGCCTG CCCGCAGTCA TCCTGCTGTG CTGTGCTTCA ATGGTCACGC | 1680 |
| TCCTTGTCCT CTTGCATTTT CCTCCCGTAC TCCTGGTTCC TCTTTGAAGT TCGTAGTAGA | 1740 |
| TAGCGGTAGA AGAAATAGTG AAAGCCTTTT TTTTTTTTTT TTTGAGGCGG AGTCTCGCTC | 1800 |
| TGTCCCCCAG GCTGGAGTGC AGTGGCGTGA TCTCGGCTCA CTGCAATCTC CGCCTCCTGG | 1860 |
| GTTACACCA TTCTCCTGCC TCACCCTCCC AAATAGCTAG GACTACAGGC GCCCTCCACC | 1920 |
| ACGCGCCCGG ATAATTTTTT GTATTTTTAG TAGAGACAGG GTTTCACCGT GTTAGCCAGG | 1980 |
| ATGGCCTCCA CCTCCTGACC TTGTGATCCG CCCGCCTCAG CCTCCCAAAG TGCTGGGATT | 2040 |
| ACAGGCGTGA GCCACCGCGC CCGCCCGAAA TAGTGAAAGT CTTAAAGTCT TTGATCTTTC | 2100 |
| TTATAAGTGC AGAGAAGAAA ACGCTGACAT ATGCTGCCTT CTCTTTCTGC TTCGGCTGCC | 2160 |
| TAAAAGGGAA GGGCCCCCTG TCCCATGATC ACGTGAAGT CTTGACCTTA TCAGTCATTT | 2220 |
| GGACGACTCA CCCTCCTTAT CCTGCCCCC CTTGTCTTGT ATACAATAA TATCAGCGCG | 2280 |
| CCCAGCCATT CGGGGCCACT ACCGGTCTCT GCGTCTTGAT GGTAGTGGTC CCCCAGGCCC | 2340 |
| AGCTGTTTTT TCTTTATCTC TTTGTCTTGT GTCTTTATTT CTTACAATCT CTCCTCTCCT | 2400 |
| CACAGGGGAA GAACACCCAC CCGCAAAGCC CCGTAGGGCT GGACCCTACG TTAGCCTGCC | 2460 |

| | |
|---|------|
| CTGCTCGGGG TTGGCGATGC TGGAGGTGGG CCTTGGACCA GAGAAAATGC TTTAATTAGG | 2520 |
| TGACAAGCGG GCAGAGGCCT TTGTCTCTGG CGCCGGCAGC CACGGCCCCC GCTGACGGCG | 2580 |
| TGGGAAACAG ACCCTGTTCC ACTCCGGTCT CCAGCCTTGG AATGGTTGCC TTCGTGCAGT | 2640 |
| GCAGGTCTGG AAAGTAGCAG TTTGGCACGG GACCCTAGAA TTCCCCAAAA GGAGTGAATA | 2700 |
| GGGGCTGGGA TTCTGGAATT TGAGTGTGGA CGGTGAGGCG GGGGGTGTGG GAGATCGGAG | 2760 |
| ACCCTGGTGG GCGCGGGAGC ACCTGCAGGC TGGAGGCCCT CGCGCGCTCC GGCGGCAGCC | 2820 |
| TGGCAAACAG GTTCTCCATC CCCAGGAGG ACGCGGCAGA GGGCGGACGA TCGCTCCACT | 2880 |
| CGCCGGGACC AGGTGCGGGG GCCCTGCCCC GCCGCTGGGG CGTGGCCAGG CTCGAAGCAC | 2940 |
| CCAGGTGTCG GGGGCCGACT CTAAGCCCTG GCACCGGAAG AGAGAGGGCG GCGGATTGGA | 3000 |
| CCTCCCGGCT CCAGCATTGC AACTGGGCGC TCCGTCTCCT GGTCCACGCA ATGATGCTGC | 3060 |
| GGCTGCTCAG AAGCCAGGTA GCCTGCCCTG GGTGAAGCCT TCGCGCAGGT CAATGACGGG | 3120 |
| GCGGAGGGGC AGGGCGCGGT CCCCTGCATC CCCGATCTGG GGAGCGGTGG GCCCAGGGGC | 3180 |
| CATCGCCTTA GCCCCTGGCG CTGGGGCTCG GCGCCAAGTG ACGGGCGGGG CTCCACCTTC | 3240 |
| CAGCCATCCG CCCGGCCCCG GAGGGCGGAC GCTGCGAGAC TCCCGGCCGC GCCCTCTCCT | 3300 |
| TCCTCTCCTC CCCAAGCCCT CGCTGCCAGT CCGGACAGGC TGCGCGGAGG GGAGGGCTGC | 3360 |
| CGGGCCGGAT AGCCGGACGC CTGGCGTTCC AGGGGCGGCC GGATGTGGCC TGCCTTTGCG | 3420 |
| GAGGGTGCGC TCCGGCCACG AAAAGCGGAC TGTGGATCTG CCACCTGCAA GCAGCTCGGG | 3480 |
| TAAGTGGGGA CTGCCCCACT CAGTTGTTCC TGGGACCCAG GAACAACTCC TTCAGAACCA | 3540 |
| GGAGGTGCAC CCCCAACCTC TTCTCCAGGT CTTCTAAGG CCCTAGGAAT CTCCGCCACC | 3600 |
| TCCCCAGCCA TTAATCCTCC AGGAACCAAG ATGCTCCTTC CGCTCCTGAC CCTCCAGCCT | 3660 |
| CTCTTGTTTT ACTTGAATA TCGTTTCCCA TCACCACCTC TGTGGTGGAT TTTGCGCCTC | 3720 |
| ACAGACAGGT ACTCCTGAGA AACAGGCTGG TGGAAGAGTC CAGTATCAGC GGAACCTACA | 3780 |
| GGAGGGGAGA CTCGAGATTC CTTCAGGAAA GGTGTAGGAA CCTGGACCAC TTTCTTTTTT | 3840 |
| TTTTTTTTTT TTTTTTTAAG ACAGGGTCCC TCTCTGTCGC GCAAGCTGGA GTGCAGTCAG | 3900 |
| CGGTGCTATC GCGGCTCATT GTGAGCTCCG GGGATCCTCC CGCCTTAGCA TCCGGTGTAG | 3960 |
| CTGAGACCAC AGACATGTGC CACCATGCCA AGCTAATTTT ATTTATTTTT TTTTGGAGAC | 4020 |

| | |
|--|------|
| GGAGTTTCAC TCTTGTTGCC CAGGCTGGAG TGTAATGGCA TGATCTCAGC TCACCGCAAC | 4080 |
| TCCCGCCCCC CGGGTTCAGG CGATTCTCCT GCCTCAGCCT CCCGAGTGGC TGGGATTACA | 4140 |
| GGCATGCGCC ACCATGCCCC GCTAATTTTG TATTTTAAAGT AGAGACAGGG TTTCTCCACG | 4200 |
| TTGGTCAGGC TGGTCTCGAA CTCCCAACCT CAGGTGATCC ACCCACCTTG GCCTCCCAAA | 4260 |
| GTGCTGGGAT TACAGGTGTG AGCCACCGCG CCTGGCCCAT GCCAAGCTAA TTTTAAAATT | 4320 |
| TTTTTGTAAG AGTGCTCTGT TGCCCAGGCT GATCTTGAAC TCCTGGGCTC AAGGGATCCT | 4380 |
| CCCATCTCAG CCTCCCAATA TGCTGGGATT ACAGGTGTGA GCCACAGTGC CCAGCCAAAC | 4440 |
| CATGGCTATC TTGAAAACCA CTTGTCTTCC AGTCCCCATG CCCCAGAAATT CCAAGGCTCT | 4500 |
| CATCCCTGAA ACCTAGGACT CAGGCTCTCC CTACCTCAGC CCCAGGAGTC TAAACCTTTA | 4560 |
| ACTTCCTCTT TCCCTGGGAC TAAGGAGTGC TGCACCCCAG GCGCCTCCCT TACCCACAT | 4620 |
| CCCTCCTCAG CCTCCCCTCC TCAGCCTCAG TGCATTTGCT AATTCGCCTT TCCTCCCCTG | 4680 |
| CAGCCATGTG GCTCCGGAGC CATCGTCAGC TCTGCCTGGC CTTCTGCTA GTCTGTGTCC | 4740 |
| TCTCTGTAAT CTTCTTCCTC CATATCCATC AAGACAGCTT TCCACATGGC CTAGGCCTGT | 4800 |
| CGATCCTGTG TCCAGACCGC CGCCTGGTGA CACCCCCAGT GGCCATCTTC TGCCTGCCGG | 4860 |
| GTACTGCGAT GGGCCCCAAC GCCTCCTCTT CCTGTCCCCA GCACCCTGCT TCCCTCTCCG | 4920 |
| GCACCTGGAC TGTCTACCCC AATGGCCGGT TTGGTAATCA GATGGGACAG TATGCCACGC | 4980 |
| TGCTGGCTCT GGCCCAGCTC AACGGCCGCC GGGCCTTTAT CCTGCCTGCC ATGCATGCCG | 5040 |
| CCCTGGCCCC GGTATTCCGC ATCACCCTGC CCGTGCTGGC CCCAGAAGTG GACAGCCGCA | 5100 |
| CGCCGTGGCG GGAGCTGCAG CTTACGACT GGATGTCGGA GGAGTACGCG GACTTGAGAG | 5160 |
| ATCCTTTCCT GAAGCTCTCT GGCTTCCCCT GCTCTTGGAC TTTCTTCCAC CATCTCCGGG | 5220 |
| AACAGATCCG CAGAGAGTTC ACCCTGCACG ACCACCTTCG GGAAGAGGCG CAGAGTGTGC | 5280 |
| TGGGTCAGCT CCGCCTGGGC CGCACAGGGG ACCGCCCCGG CACCTTTGTC GGCGTCCACG | 5340 |
| TGCGCCGTGG GGAATATCTG CAGGTTATGC CTCAGCGCTG GAAGGGTGTG GTGGGCGACA | 5400 |
| GCGCTACCT CCGGCAGGCC ATGGACTGGT TCCGGGCACG GCACGAAGCC CCCGTTTTCG | 5460 |
| TGGTCACCAG CAACGGCATG GAGTGGTGTA AAGAAAACAT CGACACCTCC CAGGGCGATG | 5520 |
| TGACGTTTGC TGGCGATGGA CAGGAGGCTA CACCGTGGA AGACTTTGCC CTGCTCACAC | 5580 |

[illegible]

TGAATGTTCC ACCATTGCTG GAAAAATTGA GATGGAAAAC AAACCATCTC TAGTTGGCCA 720C
 GCGTCTTGCT CTGTTACACAG TCTCTGGAAA AGCTGGGGTA GTTGGTGAGC AGAGCGGGAC 726C
 TCTGTCCAAC AAGCCCCACA GCCCCTCAAA GACTTTTTTTT TGT TTGTTT GAGCAGACAG 732C
 GCTAAAATGT GAACGTGGGG TGAGGGATCA CTGCCAAAAT GGTACAGCTT CTGGAGCAGA 738C
 ACTTTCAGG GATCCAGGGA CACTTTTTTTT TAAAGCTCAT AACTGCCAA GAGCTCCATA 744C
 TATTGGGTGT GAGTTCAGGT TGCCTCTCAC AATGAAGGAA GTTGGTCTTT GTCTGCAGGT 750C
 GGGCTGCTGA GGGTCTGGGA TCTGTTTTCT GGAAGTGTGC AGGTATAAAC ACACCCTCTG 756C
 TGCTTGTGAC AACTGGCAG GTACCGTGCT CATTGCTAAC CACTGTCTGT CCCTGAACTC 762C
 CCAGAACCAC TACATCTGGC TTTGGGCAGG TCTGAGATAA AACGATCTAA AGGTAGGCAG 768C
 ACCCTGGACC CAGCCTCAGA TCCAGGCAGG AGCAGGAGGT CTGGCCAAGG TGGACGGGGT 774C
 TGTCGAGATC TCAGGAGCCC CTTGCTGTTT TTTGGAGGGT GAAAGAAGAA ACCTTAAACA 780C
 TAGTCAGCTC TGATCACATC CCCTGTCTAC TCATCCAGAC CCCATGCCTG TAGGCTTATC 786C
 AGGGAGTTAC AGTTACAATT GTTACAGTAC TGT TCCCAAC TCAGCTGCCA CGGGTGAGAG 792C
 AGCAGGAGGT ATGAATTAAA AGTCTACAGC ACTAACCCGT GTCTCTGTAG CTTTTTTGGA 798C
 GCCAGAGCCA CTGTGTATGT GTGTGTGGGT TTGTGTGTGT GTGTGTGTGT GTGTGTGTGT 804C
 AAGAGAGTGG AGGAAAAGGT GGGGTACTTC TGAAGACTTT TATTTTTTTT TAATTAATTT 810C
 ATTTTTTTTC AGAGATCGAG TCTTGCTCTG TGGCCCAGGC TGGAGTGCAG TAGTGTGATC 816C
 TCGGCCCCACT GCAA 817C

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 365 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

| | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Met | Trp | Leu | Arg | Ser | His | Arg | Gln | Leu | Cys | Leu | Ala | Phe | Leu | Val |
| 1 | | | | 5 | | | | 10 | | | | | 15 | |

Cys Val Leu Ser Val Ile Phe Phe Leu His Ile His Gln Asp Ser Phe
 20 25 30
 Pro His Gly Leu Gly Leu Ser Ile Leu Cys Pro Asp Arg Arg Leu Val
 35 40 45
 Thr Pro Pro Val Ala Ile Phe Cys Leu Pro Gly Thr Ala Met Gly Pro
 50 55 60
 Asn Ala Ser Ser Ser Cys Pro Gln His Pro Ala Ser Leu Ser Gly Thr
 65 70 75 80
 Trp Thr Val Tyr Pro Asn Gly Arg Phe Gly Asn Gln Met Gly Gln Tyr
 85 90 95
 Ala Thr Leu Leu Ala Leu Ala Gln Leu Asn Gly Arg Arg Ala Phe Ile
 100 105 110
 Leu Pro Ala Met His Ala Ala Leu Ala Pro Val Phe Arg Ile Thr Leu
 115 120 125
 Pro Val Leu Ala Pro Glu Val Asp Ser Arg Thr Pro Trp Arg Glu Leu
 130 135 140
 Gln Leu His Asp Trp Met Ser Glu Glu Tyr Ala Asp Leu Arg Asp Pro
 145 150 155 160
 Phe Leu Lys Leu Ser Gly Phe Pro Cys Ser Trp Thr Phe Phe His His
 165 170 175
 Leu Arg Glu Gln Ile Arg Arg Glu Phe Thr Leu His Asp His Leu Arg
 180 185 190
 Glu Glu Ala Gln Ser Val Leu Gly Gln Leu Arg Leu Gly Arg Thr Gly
 195 200 205
 Asp Arg Pro Arg Thr Phe Val Gly Val His Val Arg Arg Gly Asp Tyr
 210 215 220
 Leu Gln Val Met Pro Gln Arg Trp Lys Gly Val Val Gly Asp Ser Ala
 225 230 235 240
 Tyr Leu Arg Gln Ala Met Asp Trp Phe Arg Ala Arg His Glu Ala Pro
 245 250 255
 Val Phe Val Val Thr Ser Asn Gly Met Glu Trp Cys Lys Glu Asn Ile
 260 265 270
 Asp Thr Ser Gln Gly Asp Val Thr Phe Ala Gly Asp Gly Gln Glu Ala
 275 280 285

104250-242560

| | | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Thr | Pro | Trp | Lys | Asp | Phe | Ala | Leu | Leu | Thr | Gln | Cys | Asn | His | Thr | Ile |
| 290 | | | | | | 295 | | | | | 300 | | | | |
| Met | Thr | Ile | Gly | Thr | Phe | Gly | Phe | Trp | Ala | Ala | Tyr | Leu | Ala | Gly | Gly |
| 305 | | | | | 310 | | | | 315 | | | | | 320 | |
| Asp | Thr | Val | Tyr | Leu | Ala | Asn | Phe | Thr | Leu | Pro | Asp | Ser | Glu | Phe | Leu |
| | | | | 325 | | | | | 330 | | | | | 335 | |
| Lys | Ile | Phe | Lys | Pro | Glu | Ala | Ala | Phe | Leu | Pro | Glu | Trp | Val | Gly | Ile |
| | | | 340 | | | | | 345 | | | | | 350 | | |
| Asn | Ala | Asp | Leu | Ser | Pro | Leu | Trp | Thr | Leu | Ala | Lys | Pro | | | |
| | | 355 | | | | | 360 | | | | | 365 | | | |

(2) INFORMATION FOR SEQ ID NO:7:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 3647 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: unknown
 - (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

| | |
|---|-----|
| CTGCAGAGAG CGCCACCCGG AAGCCACTTT TATAGAAGCT TTTACACACA ATGCTTGATT | 60 |
| TTTTTTTTTT TTTTCCGAGA CGGAGTCTCG CTTTGTCGCC CAGGCTGGAG TGCAGTGGCG | 120 |
| CGATCTGGGC TCACTGCAAG CTCCGCCTCC TGGGTTGACG CCATTCTCCT GCCTCAGCTT | 180 |
| CCCGAGTAGC TGGGACTACA GCGCCCCGCC ACCAAGCCTG GCTAATTTTT TTTTATTTTT | 240 |
| AGTGGAGACA GAGTTTCACC GTGTTAGCCA GGATGGTCTC GATCTCCTGA CCTCGGGATC | 300 |
| CGCCCGCCTC GGCCTCCCAA AGTGCTGGGA GTATAGGCGT GAGCCACCGC GCCTGGCCTA | 360 |
| TACTTGATTT TTAATGAAAA CATTCTTAAA TTCATATGGC TAACGCAAAT TTATTTTCTG | 420 |
| TAGGCATAAC ATCAAAAACA CCTGGCAGGA CTGCCCCATT CCCAGCACTG TCTAGTTCTC | 480 |
| CCCTAGTATC AGTGGGACTC CACTGATGCA CAGCTGTGAT CTACTAAAC TTCTCTCAA | 540 |
| ACTTTCTCCT CTCCTTAGGT CAGCAGCCCC GCCCCTGATC TATTTGGAAA TCCCCTGAAT | 600 |

Top50: 52429860

AAAAGTTGAA TATCATAAAC CAAAGCGAAC ACCCAGAAAT TCAAATTCAA CCCGTAGGTA 66
AAAAATTTCT CAAGTGACTG TAGACGTAGA TGTCTCCAGT GTCGCCTAAT AAGGTAGAAG 72
AGGCCAGTGC GATACTGTCT TTACACCCTT AACTTGGGTG CTAGAATATT TATCTTCGTC 78
ATCATTTTAT CATCCAAACT ATTTTGCATA ACTTTCATGG GTGCAGAAAA TGTTTTTTAA 84
GTGCTTGGTA AAATTAATAG TGATATTCAT TCATTCATCT CACTGAACAG GCAATAAATT 90
CCTTGACGAC AAGGGCCTTG GGGGGGGCCA CATCTTCATC TTTGGTTTAT GAGTCCTGTG 96
CGTCTTGGTA CAAGCAATAC TACTATGAGC CGGCAAGTCA GACTTATTTG GTAGGGGACC 102
AAAGGAAAGA ACATGTTTTG ATTGCTAAGA AAACATTTTG TTCTCTATCC TTTACTGGGC 108
TGGCAGGCAA AGGAAATGTT CTTATGAGCA CTCACATTGA AAACCTAAGT TCTTCACCAA 114
ATGCAGAGAC TCTGAAGGCC ACGCCGCTGC GGGCTGCCTC CACAATTCGA CCGTCTCGGC 120
GGGCCACGAG ATCCTGGCCA CGGATGCGGT GGCCGCGCCT CTGCTCGCAC GTTCCCCCGG 126
CCTCTGGACT CCCTCCCTCC CTCAATCCCT CCCTCCGGCG GCGCTCGCTG GCGGGTGGCT 132
AGGCCCAACG GCAGGAAGCC GACGCTATCC TCCGTTCCGC GCGCCGGGT CCGCCTTCCG 138
TCTGTTCTAG GGCCTGCTCC TGC CGGCAG CTGCTTTAGA AGGTCTCGAG CCTCCTGTAC 144
CTTCCCAGGG ATGAACCGGG CCTTCCCTCT GGAAGGCGAG GGTTCGGGCC ACAGTGAGCG 150
AGGGCCAGGG CGGTGGGCGC GCGCAGAGGG AAACCGGATC AGTTGAGAGA GAATCAAGAG 156
TAGCGGATGA GCGCCTTGTG GGGCGCGGCC CGGAAGCCCT CCGGCGCGGG CTGGGAGAAG 162
GAGTGGGCGG AGGCGCCGCA GGAGGCTCCC GGGGCCTGGT CCGGCGGGCT GGGCCCCGGG 168
CGCAGTGGAA GAAAGGGACG GGCGGTGCCC GGTGCGGCGT CCTGGCCAGC TCACCTTGCC 174
CTGGCGGCTC GCCCCGCCCC GCACTTGGGA GGAGCAGGGC AGGGCCCGCG GCCTTTGCAT 180
TCTGGGACCG CCCCTTCCA TTCCCGGGCC AGCGGCGAGC GGCAGCGACG GCTGGAGCCG 186
CAGCTACAGC ATGAGAGCCG GTGCCGCTCC TCCACGCCTG CGGACGCGTG GCGAGCGGAG 192
GCAGCGCTGC CTGTTGCGCG CATGGGGGCA CCGTGGGGCT CGCCGACGGC GGCGGCGGGC 198
GGGCGGCGCG GGTGGCGCCG AGGCCGGGGG CTGCCATGGA CCGTCTGTGT GCTGGCGGCC 204
GCCGGCTTGA CGTGACGGC GCTGATCACC TACGCTTGCT GGGGGCAGCT GCCGCCGCTG 210
CCCTGGGCGT CGCCAACCCC GTCGCGACCG GTGGGCGTGC TGCTGTGGTG GGAGCCCTTC 216

| | |
|--|------|
| GGGGGGCGCG ATAGCGCCCC GAGGCGCCCC CCTGACTGCC CGCTGCGCTT CAACATCAGC | 2220 |
| GGCTGCCGCC TGCTCACCGA CCGCGCGTCC TACGGAGAGG CTCAGGCCGT GCTTTTCCAC | 2280 |
| CACCGCGACC TCGTGAAGGG GCCCCCGAC TGGCCCCCGC CCTGGGGCAT CCAGGCGCAC | 2340 |
| ACTGCCGAGG AGGTGGATCT GCGCGTGTG GACTACGAGG AGGCAGCGGC GGCGGCAGAA | 2400 |
| GCCCTGGCGA CCTCCAGCCC CAGGCCCCCG GGCCAGCGCT GGGTTTGGAT GAACTTCGAG | 2460 |
| TCGCCCTCGC ACTCCCCGGG GCTGCGAAGC CTGGCAAGTA ACCTCTTCAA CTGGACGCTC | 2520 |
| TCCTACCGGG CGGACTCGGA CGTCTTTGTG CCTTATGGCT ACCTCTACCC CAGAAGCCAC | 2580 |
| CCCGGCGACC CGCCCTCAGG CCTGGCCCCG CCACTGTCCA GGAAACAGGG GCTGGTGGCA | 2640 |
| TGGGTGGTGA GCCACTGGGA CGACCGCCAG GCCCGGGTCC GCTACTACCA CCAACTGAGC | 2700 |
| CAACATGTGA CCGTGGACGT GTTCGGCCCG GGCGGGCCCG GGCAGCCGGT GCCCGAAATT | 2760 |
| GGGCTCCTGC ACACAGTGGC CCGCTACAAG TTCTACCTGG CTTTCGAGAA CTCGCAGCAC | 2820 |
| CTGGATTATA TCACCGAGAA GCTCTGGCGC AACGCGTTGC TCGCTGGGGC GGTGCCGGTG | 2880 |
| GTGCTGGGCC CAGACCGTGC CAACTACGAG GCGTTTGTGC CCCGCGGCGC CTTCATCCAC | 2940 |
| GTGGACGACT TCCCAAGTGC CTCCTCCCTG GCCTCGTACC TGCTTTTCCT CGACCGCAAC | 3000 |
| CCCGCGGTCT ATCGCCGCTA CTTCCACTGG CGCCGGAGCT ACGCTGTCCA CATCACCTCC | 3060 |
| TTCTGGGACG AGCCTTGGTG CCGGGTGTGC CAGGCTGTAC AGAGGGCTGG GGACCGGCCC | 3120 |
| AAGAGCATA C GGAAGTTGGC CAGCTGGTTC GAGCGGTGAA GCCGCGCTCC CCTGGAAGCG | 3180 |
| ACCCAGGGGA GCCCAAGTTG TCAGCTTTTT GATCCTCTAC TGTGCATCTC CTTGACTGCC | 3240 |
| GCATCATGGG AGTAAGTTCT TCAAACACCC ATTTTGTGCTC TATGGGAAAA AAACGATTTA | 3300 |
| CCAATTAATA TTAATCAGCA CAGAGATGGG GGCCCGGTTT CCATATTTTT TGCACAGCTA | 3360 |
| GCAATTGGGC TCCCTTTGCT GCTGATGGGC ATCATTGTTT AGGGGTGAAG GAGGGGGTTC | 3420 |
| TTCTCACCT TGTAACCAGT GCAGAAATGA AATAGCTTAG CGGCAAGAAG CCGTTGAGGC | 3480 |
| GGTTTCCTGA ATTTCCCAT CTGCCACAGG CCATATTTGT GGCCCGTGCA GCTTCCAAAT | 3540 |
| CTCATAACA ACTGTTCCCG ATTCACGTTT TTCTGGACCA AGGTGAAGCA AATTTGTGGT | 3600 |
| TGTAGAAGGA GCCTTGTTGG TGGAGAGTGG AAGGACTGTG GCTGCAG | 3660 |

(i) SEQUENCE CHARACTERISTICS:

- (ii) MOLECULE TYPE: protein

| | | | | | | | | | | | | | | | | |
|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|--|
| Met 1 | Gly | Ala | Pro | Trp 5 | Gly | Ser | Pro | Thr | Ala 10 | Ala | Ala | Gly | Gly | Arg 15 | Arg | |
| Gly | Trp | Arg | Arg 20 | Gly | Arg | Gly | Leu | Pro 25 | Trp | Thr | Val | Cys | Val 30 | Leu | Ala | |
| Ala | Ala | Gly 35 | Leu | Thr | Cys | Thr | Ala 40 | Leu | Ile | Thr | Tyr | Ala 45 | Cys | Trp | Gly | |
| Gln | Leu 50 | Pro | Pro | Leu | Pro | Trp 55 | Ala | Ser | Pro | Thr | Pro 60 | Ser | Arg | Pro | Val | |
| Gly 65 | Val | Leu | Leu | Trp | Trp 70 | Glu | Pro | Phe | Gly | Gly 75 | Arg | Asp | Ser | Ala | Pro 80 | |
| Arg | Pro | Pro | Pro | Asp 85 | Cys | Pro | Leu | Arg | Phe 90 | Asn | Ile | Ser | Gly | Cys 95 | Arg | |
| Leu | Leu | Thr | Asp 100 | Arg | Ala | Ser | Tyr | Gly 105 | Glu | Ala | Gln | Ala | Val 110 | Leu | Phe | |
| His | His | Arg 115 | Asp | Leu | Val | Lys | Gly 120 | Pro | Pro | Asp | Trp | Pro 125 | Pro | Pro | Trp | |
| Gly | Ile 130 | Gln | Ala | His | Thr | Ala 135 | Glu | Glu | Val | Asp | Leu 140 | Arg | Val | Leu | Asp | |
| Tyr 145 | Glu | Glu | Ala | Ala | Ala 150 | Ala | Ala | Glu | Ala | Leu 155 | Ala | Thr | Ser | Ser | Pro 160 | |
| Arg | Pro | Pro | Gly | Gln 165 | Arg | Trp | Val | Trp | Met 170 | Asn | Phe | Glu | Ser | Pro 175 | Ser | |
| His | Ser | Pro | Gly 180 | Leu | Arg | Ser | Leu | Ala 185 | Ser | Asn | Leu | Phe | Asn 190 | Trp | Thr | |
| Leu | Ser | Tyr 195 | Arg | Ala | Asp | Ser | Asp 200 | Val | Phe | Val | Pro | Tyr 205 | Gly | Tyr | Leu | |

Tyr Pro Arg Ser His Pro Gly Asp Pro Pro Ser Gly Leu Ala Pro Pro
 210 215 220
 Leu Ser Arg Lys Gln Gly Leu Val Ala Trp Val Val Ser His Trp Asp
 225 230 235 240
 Glu Arg Gln Ala Arg Val Arg Tyr Tyr His Gln Leu Ser Gln His Val
 245 250 255
 Thr Val Asp Val Phe Gly Arg Gly Gly Pro Gly Gln Pro Val Pro Glu
 260 265 270
 Ile Gly Leu Leu His Thr Val Ala Arg Tyr Lys Phe Tyr Leu Ala Phe
 275 280 285
 Glu Asn Ser Gln His Leu Asp Tyr Ile Thr Glu Lys Leu Trp Arg Asn
 290 295 300
 Ala Leu Leu Ala Gly Ala Val Pro Val Val Leu Gly Pro Asp Arg Ala
 305 310 315 320
 Asn Tyr Glu Arg Phe Val Pro Arg Gly Ala Phe Ile His Val Asp Asp
 325 330 335
 Phe Pro Ser Ala Ser Ser Leu Ala Ser Tyr Leu Leu Phe Leu Asp Arg
 340 345 350
 Asn Pro Ala Val Tyr Arg Arg Tyr Phe His Trp Arg Arg Ser Tyr Ala
 355 360 365
 Val His Ile Thr Ser Phe Trp Asp Glu Pro Trp Cys Arg Val Cys Gln
 370 375 380
 Ala Val Gln Arg Ala Gly Asp Arg Pro Lys Ser Ile Arg Asn Leu Ala
 385 390 395 400
 Ser Trp Phe Glu Arg
 405

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1488 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

Top250-SeqID: 9

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

ATGGGGGCAC CGTGGGGCTC GCCGACGGCG GCGGCGGGCG GGCGGCGCGG GTGGCGCCGA 60
GGCCCGGGGC TGCCATGGAC CGTCTGTGTG CTGGCGGCCG CCGGCTTGAC GTGTACGGCG 120
CTGATCACCT ACGCTTGCTG GGGGCAGCTG CCGCCGCTGC CCTGGGCGTC GCCAACCCCG 180
TCGCGACCGG TGGGCGTGCT GCTGTGGTGG GAGCCCTTCG GGGGGCGCGA TAGCGCCCCG 240
AGGCCGCCCC CTGACTGCTG CTGGGGGCAG CTGCCGCCGC TGCCCTGGGC GTCGCCAACC 300
CCGTCGCGAC CGGTGGGCGT GCTGCTGTGG TGGGAGCCCT TCGGGGGGCG CGATAGCGCC 360
CCGAGGCCGC CCCCTGACTG CCCGCTGCGC TTCAACATCA GCGGCTGCCG CCTGCTCACC 420
GACCGCGCGT CCTACGGAGA GGCTCAGGCC GTGCTTTTCC ACCACCGCGA CCTCGTGAAG 480
GGGCCCCCGG ACTGGCCCCC GCCCTGGGGC ATCCAGGCGC AACTGCCGA GCCGCTGCGC 540
TTCAACATCA GCGGCTGCCG CCTGCTCACC GACCGCGCGT CCTACGGAGA GGCTCAGGCC 600
GTGCTTTTCC ACCACCGCGA CCTCGTGAAG GGGCCCCCGG ACTGGCCCCC GCCCTGGGGC 660
ATCCAGGCGC AACTGCCGA GGAGGTGGAT CTGCGCGTGT TGGACTACGA GGAGGCAGCG 720
GCGGCGGCAG AAGCCCTGGC GACCTCCAGC CCCAGGCCCC CGGGCCAGCG CTGGGTTTGG 780
ATGAACTTCG AGTCGCCCTC GCACTCCCCG GGGCTGCGAA GCCTGGCAAG TAACCTCTTC 840
AACTGGACGC TCTCCTACCG GGCGGACTCG GACGTCTTTG TGCCTTATGG CTACCTCTAC 900
CCCAGAAGCC ACCCCGGCGA CCCGCCCTCA GGCCTGGCCC CGCCACTGTC CAGGAAACAG 960
GGGCTGGTGG CATGGGTGGT GAGCCACTGG GACGAGCGCC AGGCCCGGGT CCGCTACTAC 102
CACCAACTGA GCCAACATGT GACCGTGGAC GTGTTCGGCC GGGGCGGGCC GGGGCAGCCG 108
GTGCCCCAAA TTGGGCTCCT GCACACAGTG GCCCGCTACA AGTTCTACCT GGCTTTTCGAG 114
AACTCGCAGC ACCTGGATTA TATCACCGAG AAGCTCTGGC GCAACGCGTT GCTCGCTGGG 120
GCGGTGCCGG TGGTGCTGGG CCCAGACCGT GCCAACTACG AGCGCTTTGT GCCCCGCGGC 126
GCCTTCATCC ACGTGGACGA CTTCCCAAGT GCCTCCTCCC TGGCCTCGTA CCTGCTTTTC 132
CTCGACCGCA ACCCCGCGGT CTATCGCCGC TACTTCCACT GGCGCCGGAG CTACGCTGTC 138
CACATCACCT CCTTCTGGGA CGAGCCTTGG TGCCGGGTGT GCCAGGCTGT ACAGAGGGCT 144
GGGGACCGGC CCAAGAGCAT ACGGAACTTG GCCAGCTGGT TCGAGCGG 148

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1316 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

TTTATGACAA GCTGTGTCAT AAATTATAAC AGCTTCTCTC AGGACACTGT GGCCAGGAAG 60
TGGGTGATCT TCCTTAATGA CCCTCACTCC TCTCTCCTCT CTTCCCAGCT ACTCTGACCC 120
ATGGATCCCC TGGGCCCAGC CAAGCCACAG TGGCTGTGGC GCCGCTGTCT GGCCGGGCTG 180
CTGTTTCAGC TGCTGGTGGC TGTGTGTTTC TTCTCCTACC TGCGTGTGTC CCGAGACGAT 240
GCCACTGGAT CCCCTAGGCC AGGGCTTATG GCAGTGGAAC CTGTCACCGG GGCTCCCAAT 300
GGGTCCCGCT GCCAGGACAG CATGGCGACC CCTGCCCACC CCACCCTACT GATCCTGCTG 360
TGGACGTGGC CTTTTAACAC ACCCGTGGCT CTGCCCCGCT GCTCAGAGAT GGTGCCCCGGC 420
GCGGCCGACT GCAACATCAC TGCCGACTCC AGTGTGTACC CACAGGCAGA CGCGGTCATC 480
GTGCACCACT GGGATATCAT GTACAACCCC AGTGCCAACC TCCCGCCCCC CACCAGGCCG 540
CAGGGGCAGC GCTGGATCTG GTTCAGCATG GAGTCCCCCA GCAACTGCCG GCACCTGGAA 600
GCCCTGGACG GATACTTCAA TCTCACCATG TCCTACCGCA GCGACTCCGA CATCTTCACG 660
CCCTACGGCT GGCTGGAGCC GTGGTCCGGC CAGCCTGCCC ACCCACCCTG CAACCTCTCG 720
GCCAAGACCG AGCTGGTGGC CTGGGCGGTG TCCAAGTGA AGCCGGACTC GGCCAGGGTG 780
CGCTACTACC AGAGCCTGCA GGCTCATCTC AAGGTGGACG TGTACGGACG CTCCCACAAG 840
CCCCTGCCCA AGGGGACCAT GATGGAGACG CTGTCCCGGT ACAAGTTCTA TCTGGCCTTC 900
GAGAACTCCT TGCACCCCGA CTACATCACC GAGAAGCTGT GGAGGAACGC CCTGGAGGCC 960
TGGGCCGTGC CCGTGGTGCT GGGCCCCAGC AGAAGCAACT ACGAGAGGTT CCTGCCGCCC 1020
GACGCCTTCA TCCACGTGGA TGACTTCCAG AGCCCCAAGG ACCTGGCCCC GTACCTGCAG 1080
GAGCTGGACA AGGACCACGC CCGCTACCTG AGCTACTTTC GCTGGCGGGA GACGCTGCGG 1140

CCTCGCTCCT TCAGCTGGGC ACTGGCTTTC TGCAAGGCCT GCTGGAAGCT GCAGCAGGAA 1200
TCCAGGTACC AGACGGTGCG CAGCATAGCG GCTTGTTTCA CCTGAGAGGC CGGCATGGGG 1260
CCTGGGCTGC CAGGGACCTC ACTTTCCCAG GGCCTCACCT ACCTAGGGTC TCTAGA 1316

(2) INFORMATION FOR SEQ ID NO:11:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 374 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

| | | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Met | Asp | Pro | Leu | Gly | Pro | Ala | Lys | Pro | Gln | Trp | Leu | Trp | Arg | Arg | Cys |
| 1 | | | | 5 | | | | | 10 | | | | | 15 | |
| Leu | Ala | Gly | Leu | Leu | Phe | Gln | Leu | Leu | Val | Ala | Val | Cys | Phe | Phe | Ser |
| | | | 20 | | | | | 25 | | | | | 30 | | |
| Tyr | Leu | Arg | Val | Ser | Arg | Asp | Asp | Ala | Thr | Gly | Ser | Pro | Arg | Pro | Gly |
| | | 35 | | | | | 40 | | | | | 45 | | | |
| Leu | Met | Ala | Val | Glu | Pro | Val | Thr | Gly | Ala | Pro | Asn | Gly | Ser | Arg | Cys |
| | 50 | | | | | 55 | | | | | 60 | | | | |
| Gln | Asp | Ser | Met | Ala | Thr | Pro | Ala | His | Pro | Thr | Leu | Leu | Ile | Leu | Leu |
| 65 | | | | | 70 | | | | | 75 | | | | | 80 |
| Trp | Thr | Trp | Pro | Phe | Asn | Thr | Pro | Val | Ala | Leu | Pro | Arg | Cys | Ser | Glu |
| | | | | 85 | | | | | 90 | | | | | 95 | |
| Met | Val | Pro | Gly | Ala | Ala | Asp | Cys | Asn | Ile | Thr | Ala | Asp | Ser | Ser | Val |
| | | | 100 | | | | | 105 | | | | | 110 | | |
| Tyr | Pro | Gln | Ala | Asp | Ala | Val | Ile | Val | His | His | Trp | Asp | Ile | Met | Tyr |
| | | 115 | | | | | 120 | | | | | 125 | | | |
| Asn | Pro | Ser | Ala | Asn | Leu | Pro | Pro | Pro | Thr | Arg | Pro | Gln | Gly | Gln | Arg |
| | 130 | | | | | 135 | | | | | 140 | | | | |
| Trp | Ile | Trp | Phe | Ser | Met | Glu | Ser | Pro | Ser | Asn | Cys | Arg | His | Leu | Glu |
| 145 | | | | | 150 | | | | | 155 | | | | | 160 |
| Ala | Leu | Asp | Gly | Tyr | Phe | Asn | Leu | Thr | Met | Ser | Tyr | Arg | Ser | Asp | Ser |
| | | | | 165 | | | | | 170 | | | | | 175 | |

Asp Ile Phe Thr Pro Tyr Gly Trp Leu Glu Pro Trp Ser Gly Gln Pro
180 185 190

Ala His Pro Pro Leu Asn Leu Ser Ala Lys Thr Glu Leu Val Ala Trp
195 200 205

Ala Val Ser Asn Trp Lys Pro Asp Ser Ala Arg Val Arg Tyr Tyr Gln
210 215 220

Ser Leu Gln Ala His Leu Lys Val Asp Val Tyr Gly Arg Ser His Lys
225 230 235 240

Pro Leu Pro Lys Gly Thr Met Met Glu Thr Leu Ser Arg Tyr Lys Phe
245 250 255

Tyr Leu Ala Phe Gln Asn Ser Leu His Pro Asp Tyr Ile Thr Glu Lys
260 265 270

Leu Trp Arg Asn Ala Leu Glu Ala Trp Ala Val Pro Val Val Leu Gly
275 280 285

Pro Ser Arg Ser Asn Tyr Glu Arg Phe Leu Pro Pro Asp Ala Phe Ile
290 295 300

His Val Asp Asp Phe Gln Ser Pro Lys Asp Leu Ala Arg Tyr Leu Gln
305 310 315 320

Glu Leu Asp Lys Asp His Ala Arg Tyr Leu Ser Tyr Phe Arg Trp Arg
325 330 335

Glu Thr Leu Arg Pro Arg Ser Phe Ser Trp Ala Leu Ala Phe Cys Lys
340 345 350

Ala Cys Trp Lys Leu Gln Gln Glu Ser Arg Tyr Gln Thr Val Arg Ser
355 360 365

Ile Ala Ala Trp Phe Thr
370

(2) INFORMATION FOR SEQ ID NO:12:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1086 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: unknown
 - (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

ATGGATCCCC TGGGTGCAGC CAAGCCACAA TGGCCATGGC GCCGCTGTCT GGCCGCACTG 60
CTATTTTCAGC TGCTGGTGGC TGTGTGTTTC TTCTCCTACC TGCCTGTGTC CCGAGACGAT 120
GCCACTGGAT CCCCTAGGGC TCCCAGTGGG TCCTCCCGAC AGGACACCAC TCCCACCCGC 180
CCCACCCTCC TGATCCTGCT ATGGACATGG CCTTTCCACA TCCCTGTGGC TCTGTCCCGC 240
TGTTTCAGAGA TGGTGCCCGG CACAGCCGAC TGCCACATCA CTGCCGACCG CAAGGTGTAC 300
CCACAGGCAG ACACGGTCAT CGTGCACCAC TGGGATATCA TGTCCAACCC TAAGTCACGC 360
CTCCCACCTT CCCCAGAGGC GCAGGGGCAG CGCTGGATCT GGTTCAACTT GGAGCCACCC 420
CCTAACTGCC AGCACCTGGA AGCCCTGGAC AGATACTTCA ATCTCACCAT GTCCTACCGC 480
AGCGACTCCG ACATCTTCAC GCCCTACGGC TGGCTGGAGC CGTGGTCCGG CCAGCCTGCC 540
CACCCACCGC TCAACCTCTC GGCCAAGACC GAGCTGGTGG CCTGGGCGGT GTCCAAGTGG 600
AAGCCGGACT CAGCCAGGGT GCGCTACTAC CAGAGCCTGC AGGCTCATCT CAAGGTGGAC 660
GTGTACGGAC GCTCCCACAA GCCCCTGCCC AAGGGGACCA TGATGGAGAC GCTGTCCCGG 720
TACAAGTTCT ACCTGGCCTT CGAGAACTCC TTGCACCCCG ACTACATCAC CGAGAAGCTG 780
TGGAGGAACG CCCTGGAGGC CTGGGCCGTG CCCGTGGTGC TGGGCCCCAG CAGAAGCAAC 840
TACGAGAGGT TCCTGCCACC CGACGCCTTC ATCCACGTGG ACGACTTCCA GAGCCCCAAG 900
GACCTGGCCC GGTACCTGCA GGAGCTGGAC AAGGACCACG CCCGCTACCT GAGCTACTTT 960
CGCTGGCGGG AGACGCTGCG GCCTCGCTCC TTCAGCTGGG CACTGGATTT CTGCAAGGCC 1020
TGCTGGAAAC TGCAGCAGGA ATCCAGGTAC CAGACGGTGC GCAGCATAGC GGCTTGGTTC 1080
ACCTGA 1080

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1654 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

TTTTCTCATC TGTGAAACAG GAATAATAAC AGCTCTTCTC AGGACTCATG GCCTGGAGCT 60
TTGGTAAGCA GGAGATTGTC ATCAATGACC CTCACTCCTC TCTCCCCACT TCCCAGAGAC 120
TCTGACCCAT GGATCCCCTG GGCCCGGCCA AGCCACAGTG GTCGTGGCGC TGCTGTCTGA 180
CCACGCTGCT GTTTCAGCTG CTGATGGCTG TGTGTTTCTT CTCCTATCTG CGTGTGTCTC 240
AAGACGATCC CACTGTGTAC CCTAATGGGT CCCGCTTCCC AGACAGCACA GGGACCCCCG 300
CCCACTCCAT CCCCCTGATC CTGCTGTGGA CGTGGCCTTT TAACAAACCC ATAGCTCTGC 360
CCCGCTGCTC AGAGATGGTG CCTGGCACGG CTGACTGCAA CATCACTGCC GACCGCAAGG 420
TGTATCCACA GGCAGACGCG GTCATCGTGC ACCACCGAGA GGTCAATGTAC AACCCAGTG 480
CCCAGCTCCC ACGCTCCCCG AGGCGGCAGG GGCAGCGATG GATCTGGTTC AGCATGGAGT 540
CCCCAAGCCA CTGCTGGCAG CTGAAAGCCA TGGACGGATA CTTCAATCTC ACCATGTCCT 600
ACCGCAGCGA CTCCGACATC TTCACGCCCT ACGGCTGGCT GGAGCCGTGG TCCGGCCAGC 660
CTGCCCACCC ACCGCTCAAC CTCTCGGCCA AGACCGAGCT GGTGGCCTGG GCAGTGTCCA 720
ACTGGGGGCC AAACCTCCGCC AGGGTGGCT ACTACCAGAG CCTGCAGGCC CATCTCAAGG 780
TGGACGTGTA CGGACGCTCC CACAAGCCCC TGCCCCAGGG AACCATGATG GAGACGCTGT 840
CCCGGTACAA GTTCTATCTG GCCTTCGAGA ACTCCTTGCA CCCCAGTAC ATCACCGAGA 900
AGCTGTGGAG GAACGCCCTG GAGGCCTGGG CCGTGCCCGT GGTGCTGGGC CCCAGCAGAA 960
GCAACTACGA GAGGTTCTCTG CCACCCGACG CCTTCATCCA CGTGGACGAC TTCCAGAGCC 1020
CCAAGGACCT GGCCCGGTAC CTGCAGGAGC TGGACAAGGA CCACGCCCGC TACCTGAGCT 1080
ACTTTCGCTG GCGGGAGACG CTGCGGCCTC GCTCCTTCAG CTGGGCACTC GCTTCTGCA 1140
AGGCCTGCTG GAACTGCAG GAGGAATCCA GGTACCAGAC ACGCGGCATA GCGGCTTGGT 1200
TCACCTGAGA GGCTGGTGTG GGGCCTGGGC TGCCAGGAAC CTCATTTTCC TGGGGCCTCA 1260
CCTGAGTGGG GGCCTCATCT ACCTAAGGAC TCGTTTGCCT GAAGCTTCAC CTGCCTGAGG 1320
ACTCACCTGC CTGGGACGGT CACCTGTTGC AGCTTCACCT GCCTGGGGAT TCACCTACCT 1380
GGGTCCTCAC TTTCCTGGGG CCTCACCTGC TGGAGTCTTC GGTGGCCAGG TATGTCCCTT 1440
ACCTGGGATT TCACATGCTG GCTTCCAGGA GCGTCCCCTG CGGAAGCCTG GCCTGCTGGG 1500

GATGTCTCCT GGGGACTTTG CCTACTGGGG ACCTCGGCTG TTGGGGACTT TACCTGCTGG 1560
 GACCTGCTCC CAGAGACCTT CCACACTGAA TCTCACCTGC TAGGAGCCTC ACCTGCTGGG 1620
 GACCTCACCC TGGAGGCACT GGGCCCTGGG AACT 1654

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 359 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

| | | | | | | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Met | Asp | Pro | Leu | Gly | Pro | Ala | Lys | Pro | Gln | Trp | Ser | Trp | Arg | Cys | Cys | 1 | 5 | 10 | 15 |
| Leu | Thr | Thr | Leu | Leu | Phe | Gln | Leu | Leu | Met | Ala | Val | Cys | Phe | Phe | Ser | 20 | 25 | 30 | |
| Tyr | Leu | Arg | Val | Ser | Gln | Asp | Asp | Pro | Thr | Val | Tyr | Pro | Asn | Gly | Ser | 35 | 40 | 45 | |
| Arg | Phe | Pro | Asp | Ser | Thr | Gly | Thr | Pro | Ala | His | Ser | Ile | Pro | Leu | Ile | 50 | 55 | 60 | |
| Leu | Leu | Trp | Thr | Trp | Pro | Phe | Asn | Lys | Pro | Ile | Ala | Leu | Pro | Arg | Cys | 65 | 70 | 75 | 80 |
| Ser | Glu | Met | Val | Pro | Gly | Thr | Ala | Asp | Cys | Asn | Ile | Thr | Ala | Asp | Arg | 85 | 90 | 95 | |
| Lys | Val | Tyr | Pro | Gln | Ala | Asp | Ala | Val | Ile | Val | His | His | Arg | Glu | Val | 100 | 105 | 110 | |
| Met | Tyr | Asn | Pro | Ser | Ala | Gln | Leu | Pro | Arg | Ser | Pro | Arg | Arg | Gln | Gly | 115 | 120 | 125 | |
| Gln | Arg | Trp | Ile | Trp | Phe | Ser | Met | Glu | Ser | Pro | Ser | His | Cys | Trp | Gln | 130 | 135 | 140 | |
| Leu | Lys | Ala | Met | Asp | Gly | Tyr | Phe | Asn | Leu | Thr | Met | Ser | Tyr | Arg | Ser | 145 | 150 | 155 | 160 |
| Asp | Ser | Asp | Ile | Phe | Thr | Pro | Tyr | Gly | Trp | Leu | Glu | Pro | Trp | Ser | Gly | 165 | 170 | 175 | |

Gln Pro Ala His Pro Pro Leu Asn Leu Ser Ala Lys Thr Glu Leu Val
180 185 190

Ala Trp Ala Val Ser Asn Trp Gly Pro Asn Ser Ala Arg Val Arg Tyr
195 200 205

Tyr Gln Ser Leu Gln Ala His Leu Lys Val Asp Val Tyr Gly Arg Ser
210 215 220

His Lys Pro Leu Pro Gln Gly Thr Met Met Glu Thr Leu Ser Arg Tyr
225 230 235 240

Lys Phe Tyr Leu Ala Phe Glu Asn Ser Leu His Pro Asp Tyr Ile Thr
245 250 255

Glu Lys Leu Trp Arg Asn Ala Leu Glu Ala Trp Ala Val Pro Val Val
260 265 270

Leu Gly Pro Ser Arg Ser Asn Tyr Glu Arg Phe Leu Pro Pro Asp Ala
275 280 285

Phe Ile His Val Asp Asp Phe Gln Ser Pro Lys Asp Leu Ala Arg Tyr
290 295 300

Leu Gln Glu Leu Asp Lys Asp His Ala Arg Tyr Leu Ser Tyr Phe Arg
305 310 315 320

Trp Arg Glu Thr Leu Arg Pro Arg Ser Phe Ser Trp Ala Leu Ala Phe
325 330 335

Cys Lys Ala Cys Trp Lys Leu Gln Glu Glu Ser Arg Tyr Gln Thr Arg
340 345 350

Gly Ile Ala Ala Trp Phe Thr
355

0005475-0544